

University of Groningen

Visualizing High-Dimensional Chemical Abundance Space in GALAH DR2

Kim, Youngjoo; Trager, Scott; Telea, Alexandru; Roerdink, J.B.T.M.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Final author's version (accepted by publisher, after peer review)

Publication date:

2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Kim, Y., Trager, S., Telea, A., & Roerdink, J. B. T. M. (2019). *Visualizing High-Dimensional Chemical Abundance Space in GALAH DR2*. Poster session presented at the Astronomical Data Analysis and Software Systems conference (ADASS) 2019, Groningen, Netherlands.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Visualizing High-Dimensional Chemical Abundance Space in GALAH DR2

Youngjoo Kim^{1, 2}

Scott C. Trager²

Alexandru C. Telea¹

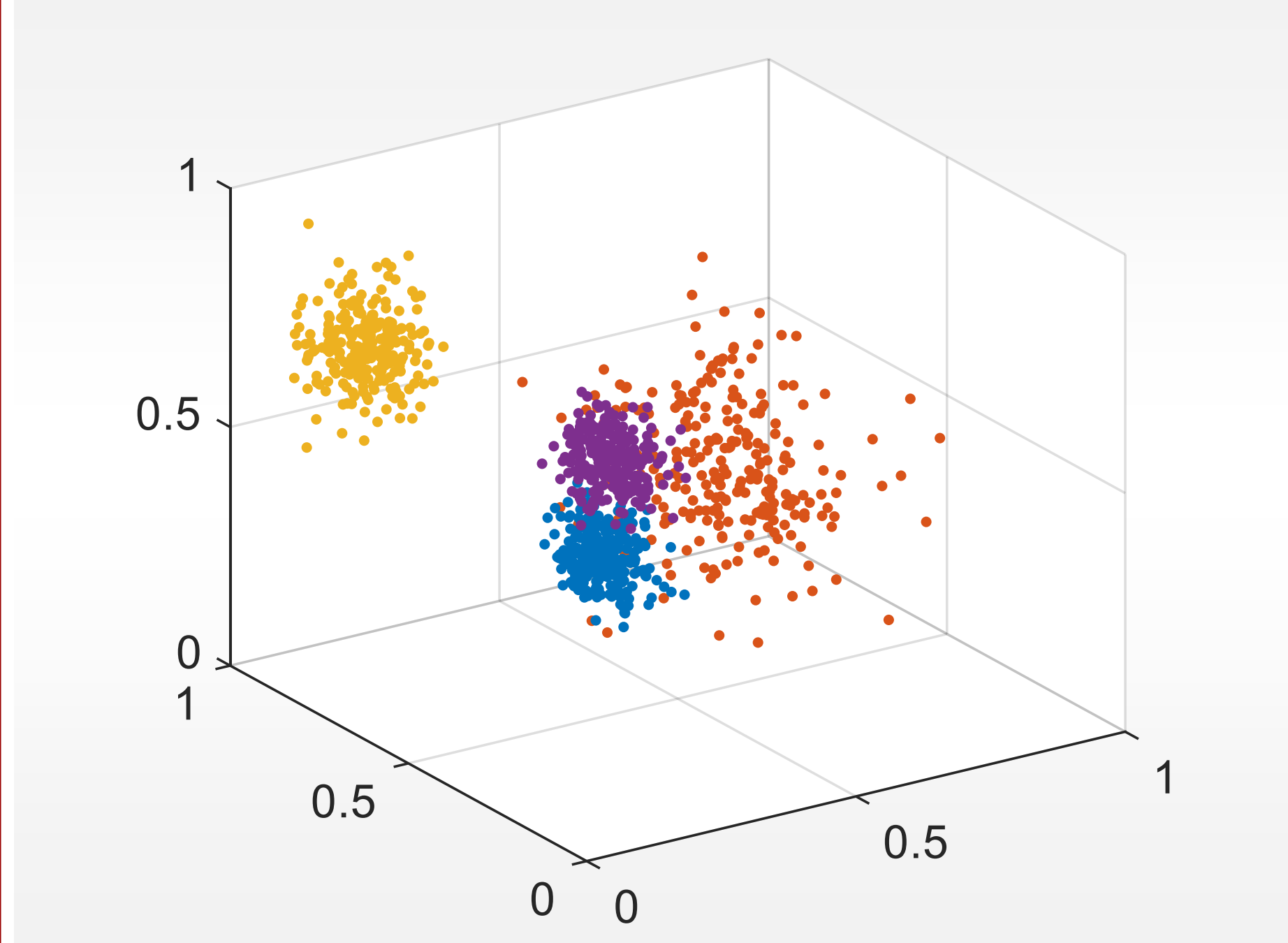
Jos B. T. M. Roerdink¹

¹Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen

²Kapteyn Astronomical Institute, University of Groningen

Aim: Visualize high-dimensional data to find interesting patterns and underlying structures

High-Dimensional Data



Gaussian random data with four clusters in 3D (also applicable to nD)

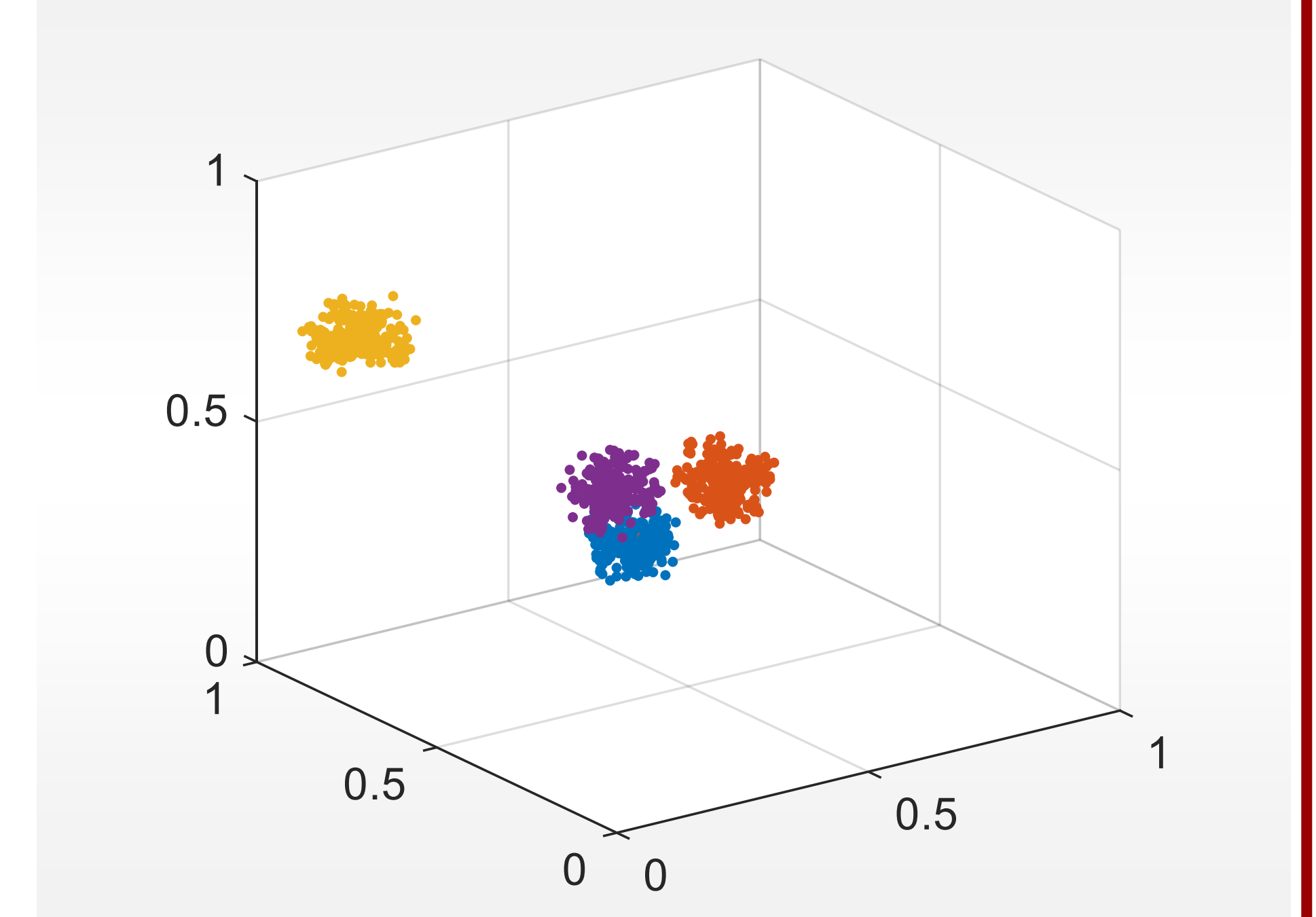
Filter high-dimensional data

“Whoop”

Local Gradient Clustering (LGC)

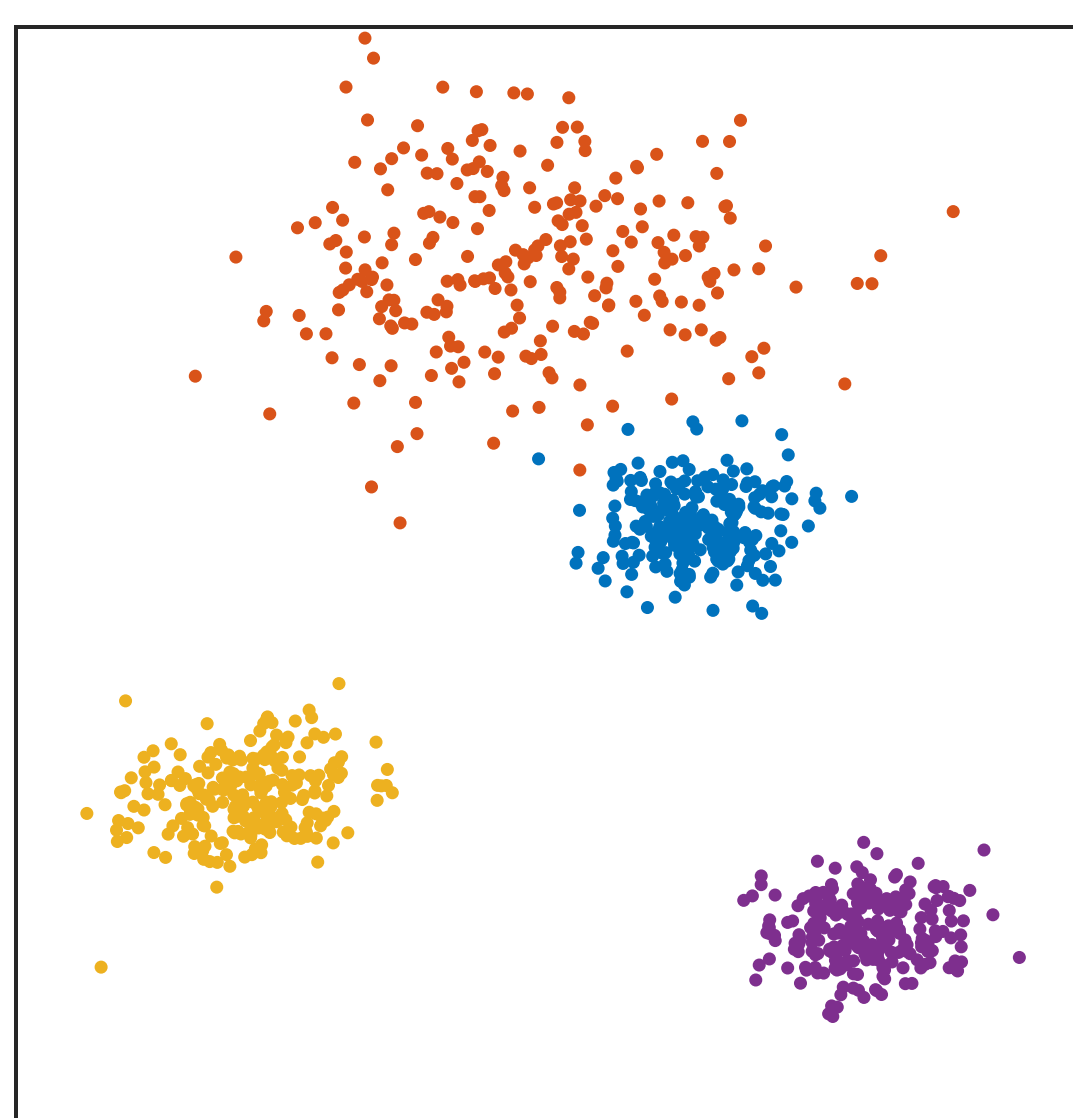
Shift points along the gradient of the kernel density estimator

Filtered Data

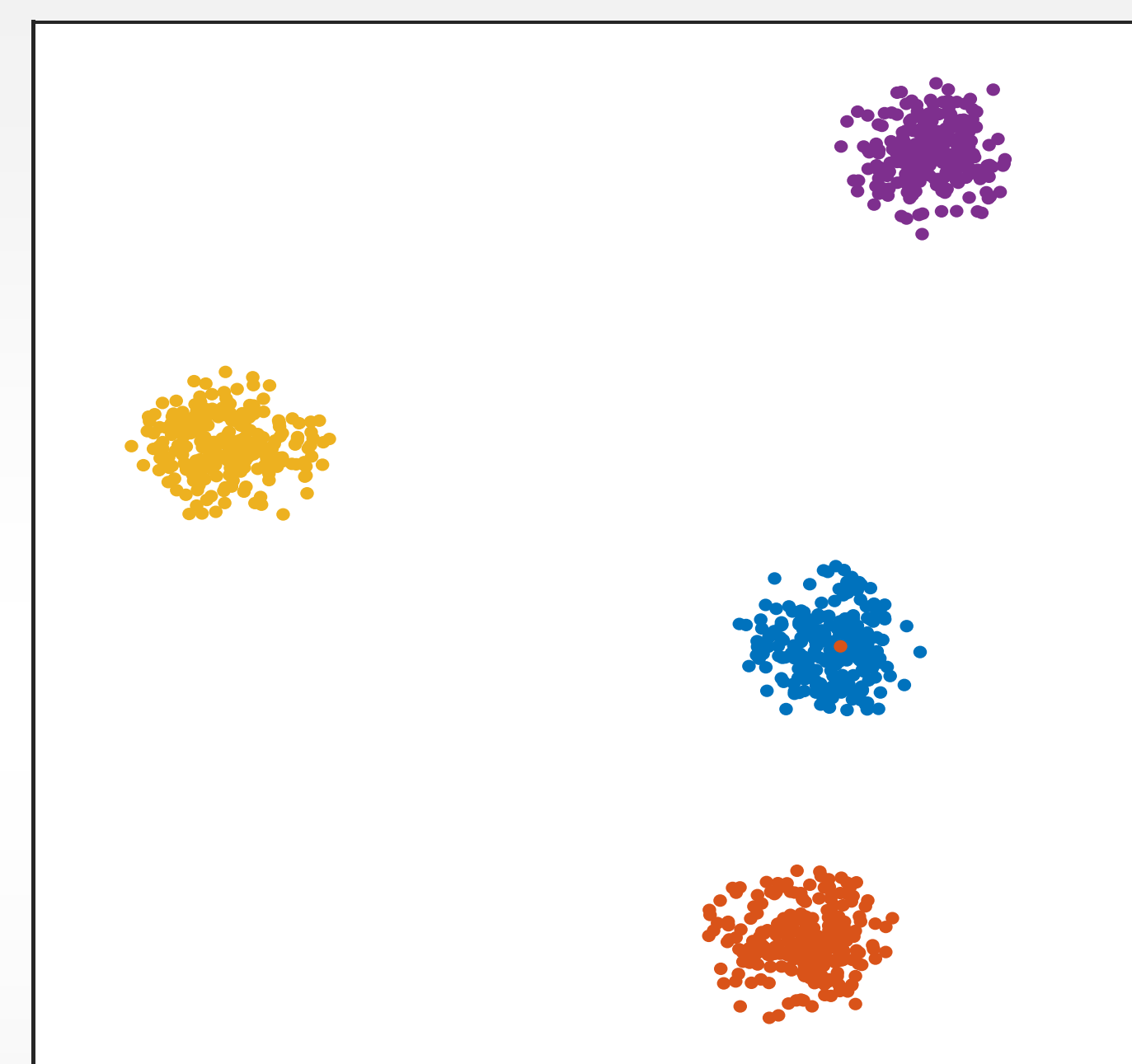


Clusters are separated in 3D

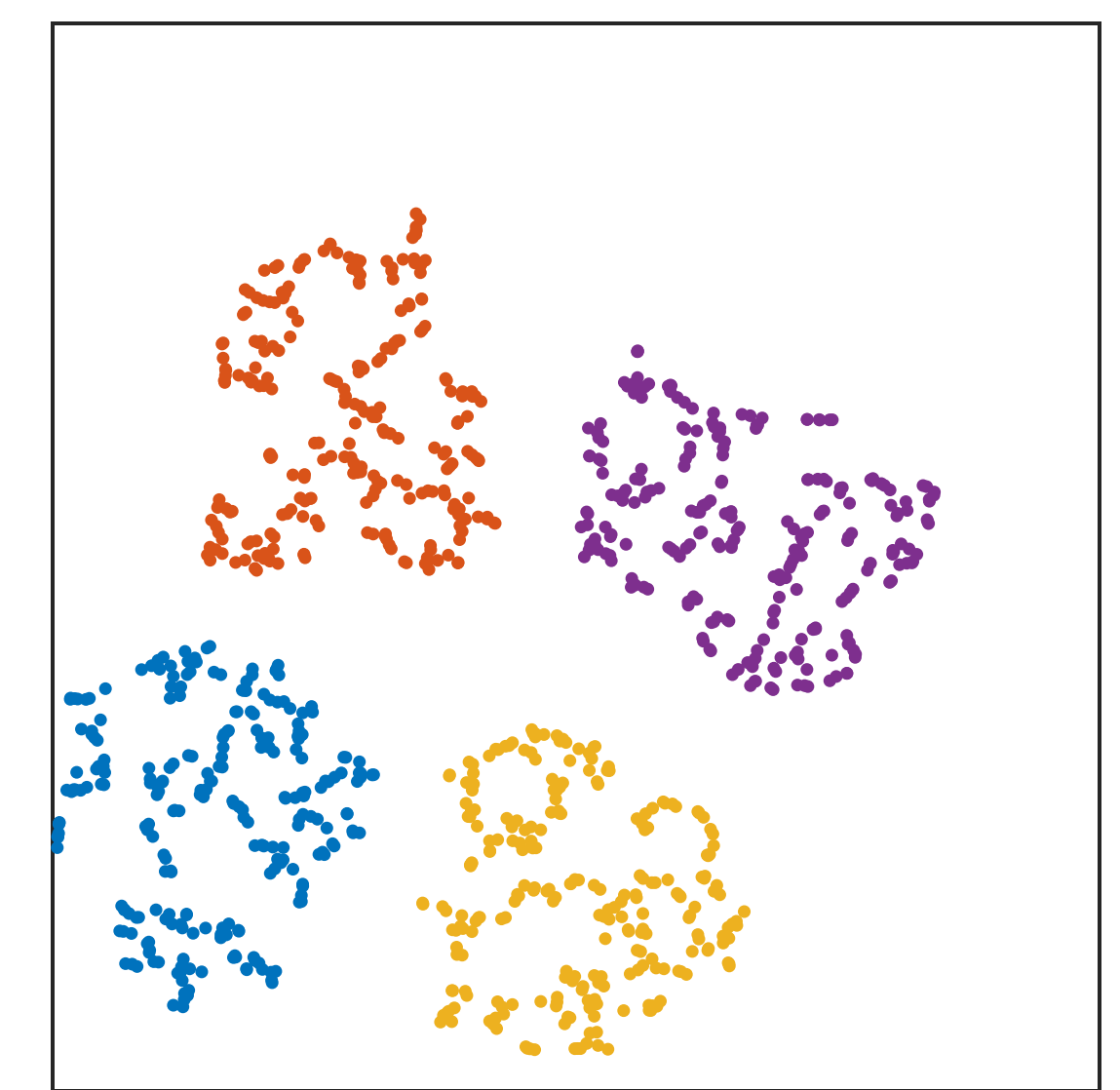
Dimensionality Reduction



VS.



VS.



Landmark Multidimensional Scaling (LMDS [1]):

Clusters are **not well separated**.
Method is **fast**.

Proposed method (LGC+LMDS):

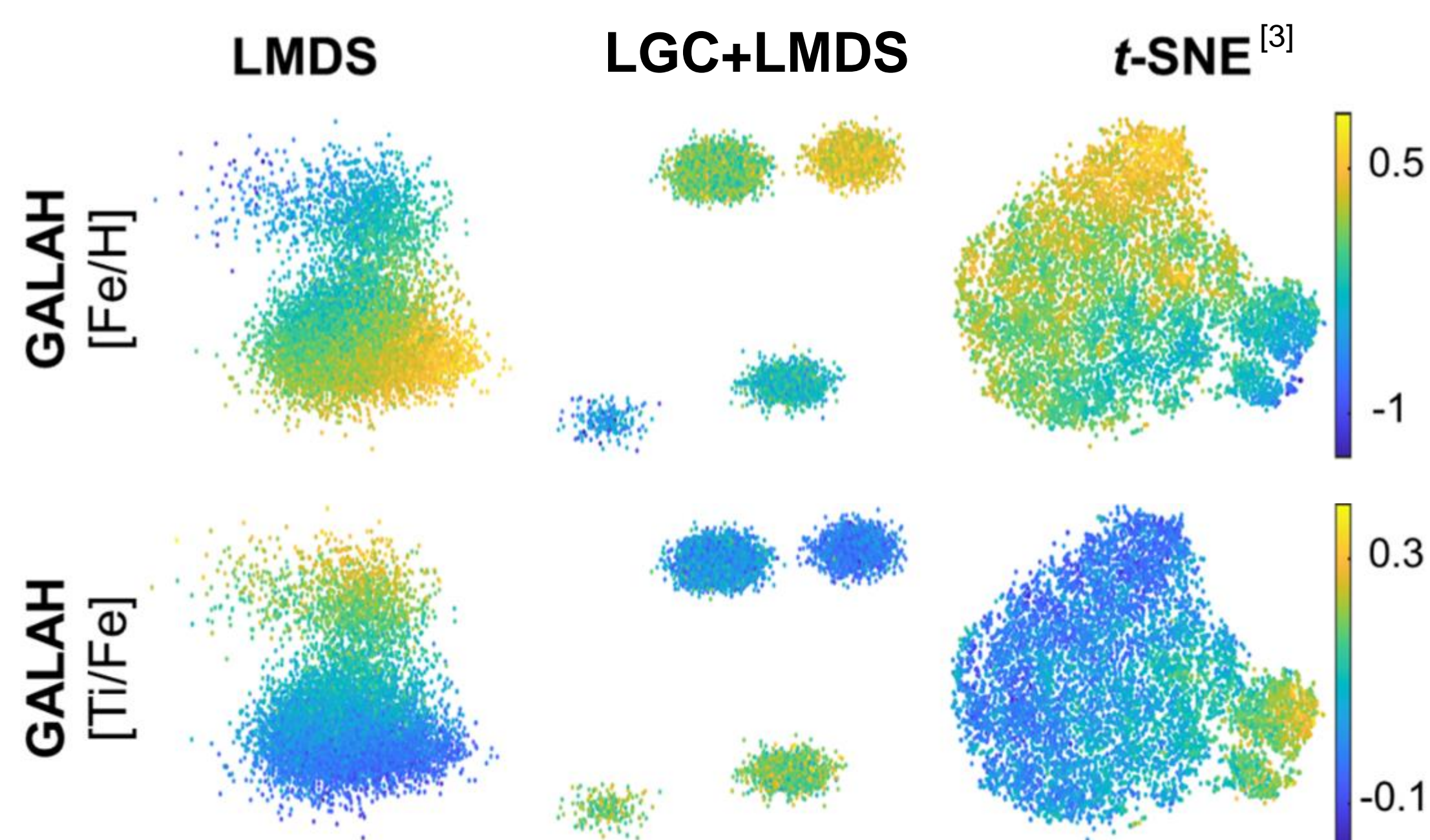
Clusters are **well separated** in the 2D projection.
Method is **fast**.

t-Stochastic Neighbor Embedding (t-SNE [2]):

Clusters are **well separated**.
Method is **slow**.

GALAH DR2

- Dataset:** 10K observations are randomly chosen from the second data release of GALactic Archaeology with HERMES survey (GALAH DR2) [4] cross-matched with *Gaia* DR2 [5-6]. 10-D data set that consists of the following 10 stellar abundances are used: [Fe/H], [Mg/Fe], [Al/Fe], [Si/Fe], [Ca/Fe], [Ti/Fe], [Cu/Fe], [Zn/Fe], [Y/Fe], and [Ba/Fe]
- Results:** LGC+LMDS shows cleaner separation of substructures in the 2D abundance-space than the original LMDS and t-SNE



References

- [1] V. De Silva and J. B. Tenenbaum, “Sparse multidimensional scaling using landmark points,” Technical report, Stanford University, Vol. 120, 2004.
- [2] L. V. D. Maaten and G. Hinton, “Visualizing data using t-SNE,” Journal of machine learning research, No. 9, pp. 2579-2605, 2008.
- [3] F. Anders *et al.*, “Dissecting stellar chemical abundance space with t-sne,” Astronomy & Astrophysics, Vol. 619, No. A125, 2018.
- [4] S. Buder *et al.*, “The GALAH Survey: Second data release,” Monthly Notices of the Royal Astronomical Society, Vol. 478, 2018.
- [5] Gaia Collaboration, “The Gaia mission,” Astronomy & Astrophysics, Vol. 595, No. A1, 2016.
- [6] Gaia Collaboration, “Gaia Data Release 2-Summary of the contents and survey properties,” Astronomy & Astrophysics, Vol. 616, No. A1, 2018.
- [7] M. Muja and D. G. Lowe, “Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration,” International Conference on Computer Vision Theory and Applications (VISAPP'09), 2009.
- [8] V. A. Epanechnikov, “Non-parametric estimation of a multivariate probability density,” Theory of Probability and its Applications, Vol. 14, No. 1, pp. 153-158, 1969.
- [9] K. Fukunaga and L. Hostetler, “The estimation of the gradient of a density function, with applications in pattern recognition,” IEEE Transactions on information theory, Vol. 21, No. 1, pp. 32-40, 1975.

Summary

Key idea

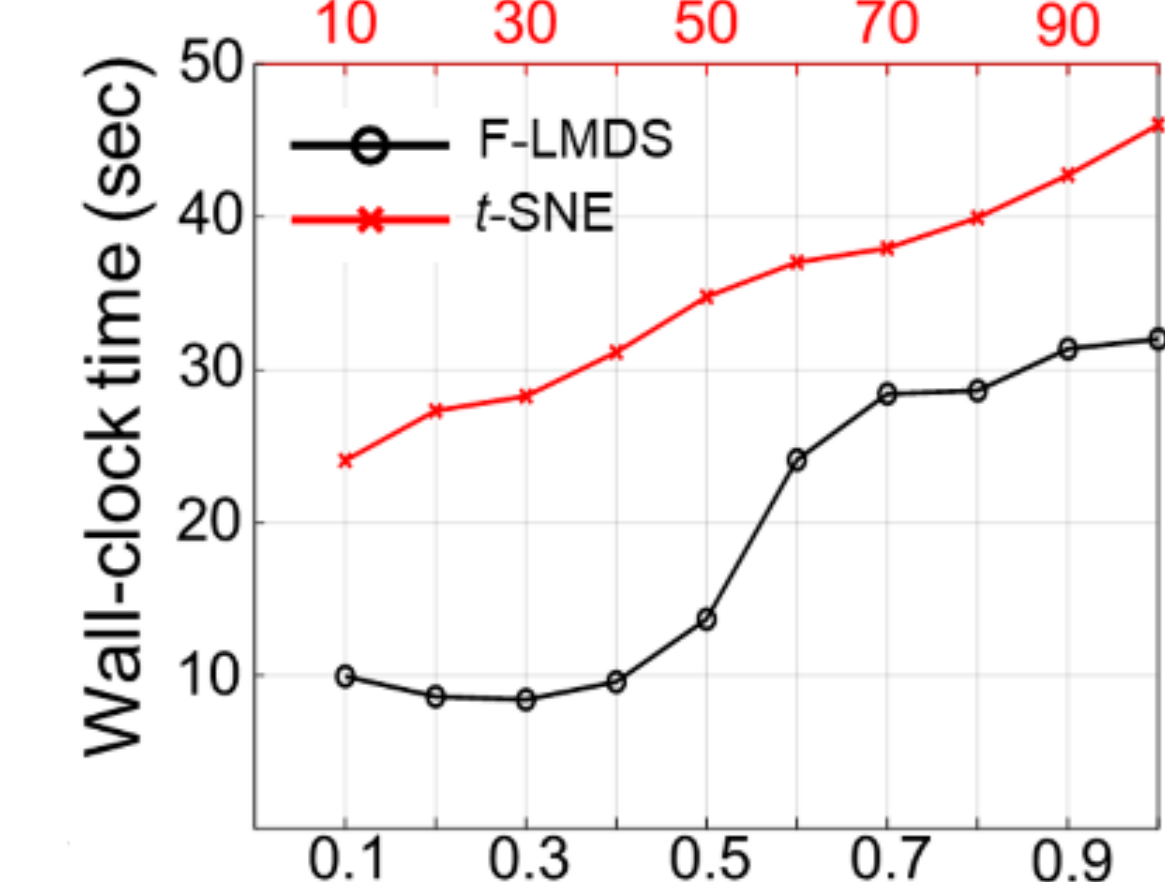
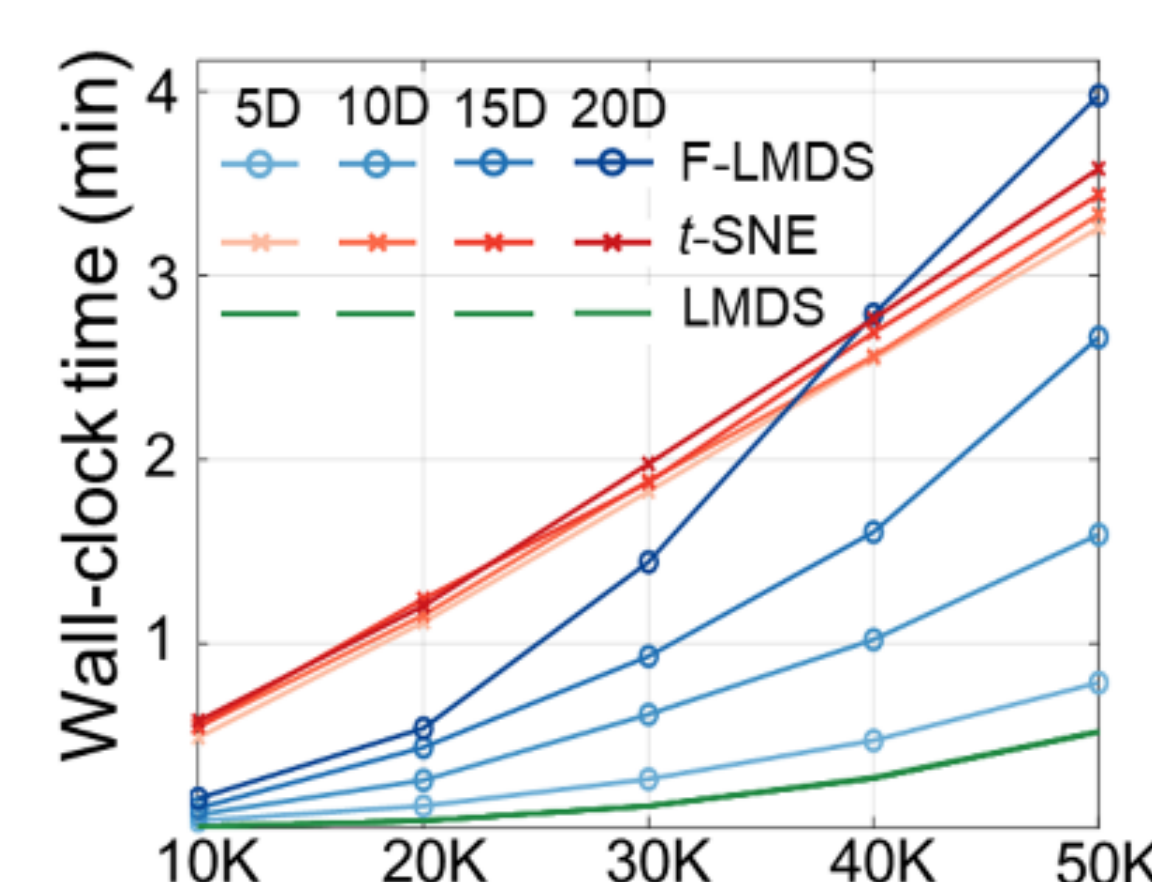
Filter the high-dimensional data so that potential clusters are well separated even after dimensionality reduction

Method

- Estimate density using Epanechnikov kernel [7-8]
- Shift points upstream in kernel density gradient, resulting in cluster contraction [9]
- Perform LMDS [1]

Advantages

- Clusters are **well separated** after the projection by preprocessing the data with local-based gradient clustering
- Predictable** outcome with one parameter
- More **computationally scalable** than t-SNE, in terms of wall-clock time



Future Work

- A more sophisticated analysis of the different substructures gained from the LGC+LMDS results using GALAH DR2